

Title (en)
ONLINE SAMPLING ANALYSIS

Title (de)
ONLINE-PROBENANALYSE

Title (fr)
ANALYSE D'ÉCHANTILLONNAGE EN LIGNE

Publication
EP 3762884 A4 20210127 (EN)

Application
EP 19763674 A 20190301

Priority
• US 201815912326 A 20180305
• US 2019020399 W 20190301

Abstract (en)
[origin: US2019272482A1] Methods, systems and computer program products generating diverse and representative set of samples from a large amount of transaction data are disclosed. A data sampling system receives transaction records. Each transaction record has multiple text segments. The system selects a subset of transaction records that contain least frequently appeared text segments. The system determines a respective vector representation for each selected transaction record. The system can measure similarity between transaction records based on the vector representations. The system assigns the selected transaction records to multiple clusters based on the vector representations and designated dimensions of importance. The system identifies one or more anchors that include transaction records on boundaries between clusters. The system filters the subset of transaction records by removing transaction records that are close to the anchors. The system then provides the filtered subset as a representative set of samples to a sample consumer.

IPC 8 full level
G06F 16/35 (2019.01); **G06N 20/00** (2019.01); **G06Q 10/06** (2012.01); **G06Q 30/02** (2012.01)

CPC (source: EP US)
G06F 16/355 (2019.01 - EP US); **G06N 3/02** (2013.01 - EP); **G06N 20/00** (2019.01 - EP US)

Citation (search report)
• [I] US 2017206272 A1 20170720 - NEELA RAGHURAM [US], et al
• [I] US 2009234683 A1 20090917 - ANDERSON RUSSELL [GB], et al
• [I] US 2016078367 A1 20160317 - ADJAOUTE AKLI [US]
• See also references of WO 2019173161A1

Designated contracting state (EPC)
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

Designated extension state (EPC)
BA ME

DOCDB simple family (publication)
US 11494687 B2 20221108; US 2019272482 A1 20190905; CA 3093075 A1 20190912; EP 3762884 A1 20210113; EP 3762884 A4 20210127; WO 2019173161 A1 20190912

DOCDB simple family (application)
US 201815912326 A 20180305; CA 3093075 A 20190301; EP 19763674 A 20190301; US 2019020399 W 20190301